Analysis of substrate specificity of Trypanosoma brucei oligosaccharyltransferases (OSTs) by functional expression of domain-swapped chimeras in yeast

Poljak, Kristina ; Breitling, Jörg ; Gauss, Robert ; Rugarabamu, George ; Pellanda, Mauro ; Aebi, Markus

Abstract: N-Linked protein glycosylation is an essential and highly conserved post-translational modification in eukaryotes. The transfer of a glycan from a lipid-linked oligosaccharide (LLO) donor to the asparagine residue of a nascent polypeptide chain is catalyzed by an oligosaccharyltransferase (OST) in the lumen of the endoplasmic reticulum (ER). Trypanosoma brucei encodes three paralogue single-protein OSTs called TbSTT3A, TbSTT3B, and TbSTT3C that can functionally complement the Saccharomyces cerevisiae OST, making it an ideal experimental system to study the fundamental properties of OST activity. We characterized the LLO and polypeptide specificity of all three TbOST isoforms and their chimeric forms in the heterologous expression host S. cerevisiae where we were able to apply yeast genetic tools and newly developed glycoproteomics methods. We demonstrated that TbSTT3A accepted LLO substrates ranging from Man5GlcNAc2 to Man7GlcNAc2 In contrast, TbSTT3B required more complex precursors ranging from Man6GlcNAc2 to Glc3Man9GlcNAc2 structures, and TbSTT3C did not display any LLO preference. Sequence differences between the isoforms cluster in three distinct regions. We have swapped the individual regions between different OST proteins and identified region 2 to influence the specificity toward the LLO and region 1 to influence polypeptide substrate specificity. These results provide a basis to further investigate the molecular mechanisms and contribution of single amino acids in OST interaction with its substrates.

DOI: https://doi.org/10.1074/jbc.M117.811133

Analysis of substrate specificity of *Trypanosoma brucei* oligosaccharyltransferases (OSTs) by functional expression of domain-swapped chimeras in yeast

Received for publication, August 9, 2017, and in revised form, October 11, 2017 Published, Papers in Press, October 17, 2017, DOI 10.1074/jbc.M117.811133

Kristina Poljak[1], **Jörg Breitling, Robert Gauss**[2], **George Rugarabamu**[3], **Mauro Pellanda**[4], **and Markus Aebi**[5]

From the Institute of Microbiology, ETH Zurich, Vladimir-Prelog-Weg 4, CH-8093 Zurich, Switzerland

Edited by Gerald W. Hart

N-Linked protein glycosylation is an essential and highly conserved post-translational modification in eukaryotes. The transfer of a glycan from a lipid-linked oligosaccharide (LLO) donor to the asparagine residue of a nascent polypeptide chain is catalyzed by an oligosaccharyltransferase (OST) in the lumen of the endoplasmic reticulum (ER). *Trypanosoma brucei* encodes three paralogue single-protein OSTs called *Tb*STT3A, *Tb*STT3B, and *Tb*STT3C that can functionally complement the *Saccharomyces cerevisiae* OST, making it an ideal experimental system to study the fundamental properties of OST activity. We characterized the LLO and polypeptide specificity of all three *Tb*OST isoforms and their chimeric forms in the heterologous expression host *S. cerevisiae* where we were able to apply yeast genetic tools and newly developed glycoproteomics methods. We demonstrated that *Tb*STT3A accepted LLO substrates ranging from $Man_5GlcNAc_2$ to $Man_7GlcNAc_2$. In contrast, *Tb*STT3B required more complex precursors ranging from $Man_6GlcNAc_2$ to $Glc_3Man_9GlcNAc_2$ structures, and *Tb*STT3C did not display any LLO preference. Sequence differences between the isoforms cluster in three distinct regions. We have swapped the individual regions between different OST proteins and identified region 2 to influence the specificity toward the LLO and region 1 to influence polypeptide substrate specificity. These results provide a basis to further investigate the molecular mechanisms and contribution of single amino acids in OST interaction with its substrates.

Asparagine-linked protein glycosylation (*N*-glycosylation) is a highly conserved post-translational modification in eukaryotes. *N*-Glycosylation is an essential process and plays important roles in protein folding quality control, cell–cell interactions, and developmental processes (1, 2). The glycans transferred to nascent polypeptide chains in the ER[6] are built on the lipid carrier dolichylphosphate (Dol-P) to yield the lipid-linked oligosaccharide (LLO) substrate for oligosaccharyltransferase (OST). The biosynthesis of the LLO is an ordered, stepwise process conducted by the concerted action of specific glycosyltransferases that are encoded by the asparagine-linked glycosylation (*ALG*) genes. LLO biosynthesis is initiated by the addition of *N*-acetylglucosaminyl-phosphate (GlcNAc-P) and *N*-acetylglucosamine (GlcNAc) to the lipid carrier on the cytosolic face of the ER membrane using nucleotide-activated UDP-GlcNAc as donor to form Dol-PP-$GlcNAc_2$. Subsequently, the LLO is elongated by five mannose (Man) residues. The $Man_5GlcNAc_2$ LLO is then translocated into the ER lumen. There, Dol-P-bound Man serves as donor for the further elongation of the LLO with Man until a $Man_9GlcNAc_2$ structure is built up. In most of the fungi and animal species, the addition of three glucose (Glc) residues from Dol-P-Glc terminates LLO biosynthesis (3).

The mature $Glc_3Man_9GlcNAc_2$ LLO is used as donor substrate by the OST that transfers the oligosaccharide from the lipid carrier *en bloc* to an asparagine residue of a nascent polypeptide chain. In eukaryotes, the acceptor asparagine residue is located within a conserved sequon consisting of three amino acids: asparagine, a second amino acid (any, but proline), and threonine or serine (N*X*(T/S)) (4). In multicellular eukaryotes the OST is a complex assembled from eight different proteins with STT3 encoding the catalytic subunit (5–7). Other subunits of the hetero-oligomeric complex were suggested to influence OST substrate interactions and complex assembly (8–12).

The genomes of the kinetoplastids *Trypanosoma* and *Leishmania* only encode homologues of the yeast *STT3* gene (13). All other subunits found in hetero-oligomeric OST complexes of *Saccharomyces cerevisiae* or mammals are missing in *Trypanosoma brucei*, *Trypanosoma cruzi*, and *Leishmania major*, suggesting that these proteins function as single subunit OSTs similar to enzymes known from bacterial and archaeal *N*-gly-

This work was supported by Swiss National Science Foundation Grants 310030B_144083 and 310030_162636 (to M. A.) and ETH Zürich. The authors declare that they have no conflicts of interest with the contents of this article.

This article contains supplemental Fig. S1 and Tables S1–S4.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (http://www.ebi.ac.uk/pride) partner repository with the dataset identifier PXD007215.

[1] Member of the Life Science Zurich Graduate Program for Molecular Life Science.

[2] Present address: Freies Gymnasium Zürich, Arbenzstrasse 19, 8008 Zürich, Switzerland.

[3] Present address: Policy Cures Research, Paramount House, 55 Brisbane St., Sydney, New South Wales 2010, Australia.

[4] Present address: Dept. of Biochemistry, University of Zurich, Winterthurerstrasse 190, Zurich 8057, Switzerland.

[5] To whom correspondence should be addressed: ETH Zürich, Institute of Microbiology, Vladimir Prelog weg 4, 8093 Zürich. Tel.: 41-44-632-64-13; Fax: 41-44-632-11-48; E-mail: markus.aebi@micro.biol.ethz.ch.

[6] The abbreviations used are: ER, endoplasmic reticulum; OST, oligosaccharyltransferase; STT3, staurosporine and temperature-sensitive 3; LLO, lipid-linked oligosaccharide; PRM, parallel reaction monitoring; SILAC, stable isotope labeling in cell culture; Dol-P, dolichylphosphate; ER, endoplasmic reticulum; VSG, variant surface glycoprotein; f-FOA, 5-fluoroorotic acid; EndoH, endoglycosidase H; L, light; H, heavy; AGC, automatic gain control.

cosylation systems (14, 15). *S. cerevisiae* has proven to be a suitable heterologous *in vivo* system to functionally express and characterize kinetoplastid OSTs. The yeast cells used in these studies lacked the essential yeast STT3 subunit or other essential OST subunits. This deleterious loss of a functional yeast OST is complemented by expression of different STT3 proteins from *T. cruzi*, *T. brucei,* and *L. major*, demonstrating that these STT3s function as single protein OSTs (16–19).

T. brucei is a protozoan parasite causing African sleeping sickness in humans and nagana in cattle. *T. brucei* cells encounter two hosts during their life cycle. The parasite exists as the procyclic form in an insect vector (the tsetse fly), and it is referred to as bloodstream form when afflicting the mammalian host. The surface of *T. brucei* cells is covered by glycoproteins termed variant surface glycoprotein (VSG) in the bloodstream form and procyclins in its procyclic life stage. *N*-Glycosylation epitopes on VSG play an important role in *T. brucei* virulence (20). Furthermore, a possible immune evasion strategy has been proposed where *T. brucei* genetically recombined its *N*-glycosylation machinery, resulting in the change of the glycosylation status of VSG (21). The *T. brucei* genome encodes three paralogues of STT3 termed *Tb*STT3A, *Tb*STT3B, and *Tb*STT3C (19, 22). In contrast to multicellular eukaryotes, trypanosomatids are incapable of synthesizing Dol-P-Glc and therefore lack the three capping Glc residues found in yeast and higher eukaryotes (22–24).

The three paralogue STT3s encoded by the *T. brucei* genome display distinct preferences for the LLO donor as well as for the acceptor polypeptide substrate. Whereas *Tb*STT3A was shown to preferentially transfer $Man_5GlcNAc_2$ glycans to acceptor polypeptide chains, both *Tb*STT3B and *Tb*STT3C glycosylated acceptor sites with $Man_9GlcNAc_2$ glycans (19). A recent study performed in *T. brucei* cells extended these findings, reporting that both *Tb*STT3A and *Tb*STT3B transfer $Man_7GlcNAc_2$ and $Man_5GlcNAc_2$ glycans to a VSG protein (25). The authors suggest that the substrate specificity of *Tb*STT3A and *Tb*STT3B is promoted by the presence or absence of the LLO substrate c-branch. The LLO specificity of *Tb*STT3C was not addressed because this paralogue is not expressed in *T. brucei* cells used for the experiments (19). Analysis of the OST protein products revealed that *Tb*STT3A and *Tb*STT3C preferentially glycosylate sequons with acidic amino acids in the sequon's vicinity. By contrast, *Tb*STT3B did not display a particular preference for glycosylation sequons and the amino acids surrounding them (19).

In this study, we used the functional expression of *T. brucei* STT3 proteins in yeast to analyze their function in detail. We took advantage of the "genetic tailoring" (26) of the OST substrate and the quantitative analysis of glycosylation site occupancy (27) to characterize *T. brucei* OST function. Domain-swap experiments made it possible to assign functional properties to specific regions of the STT3 proteins.

Results

TbSTT3A, TbSTT3B, and TbSTT3C display differential preferences for LLO substrates from $Man_5GlcNAc_2$ to $Glc_3Man_9GlcNAc_2$

In vivo analysis of the *Tb*STT3 LLO specificity revealed that *Tb*STT3B transfers primarily $Man_9GlcNAc_2$ oligosaccharides, whereas *Tb*STT3A transfers $Man_5GlcNAc_2$ glycans to its preferred glycosylation site of VSG221. Data acquired in the Δ*stt3 S. cerevisiae* strain suggested that *Tb*STT3B and *Tb*STT3C both accept $Glc_3Man_9GlcNAc_2$ as LLO substrate, whereas *Tb*STT3A cannot utilize the $Glc_3Man_9GlcNAc_2$ LLO substrate (19). We therefore further investigated the LLO specificity of all three *Tb*STT3 paralogues using the heterologous yeast expression system. We combined the *STT3* deletion with different deletions in the LLO biosynthesis pathway (*ALG* genes) in yeast. The Δ*stt3*Δ*alg* strains harboring the yeast *STT3* (*ScSTT3*) *URA3* plasmid and a second *LEU2*-marked plasmid, encoding either *ScSTT3* or the different *TbSTT3* paralogues, were subjected to plasmid shuffling using 5-FOA. In this approach, survival of the cells on 5-FOA depended on the ability of the different *TbSTT3* genes, encoded on the *LEU2* plasmids, to complement the yeast *STT3* deletion.

In the Δ*stt3*Δ*alg3* double mutant strain (accumulation of the $Man_5GlcNAc_2$ oligosaccharide), only *Tb*STT3A and *Tb*STT3C complemented the *STT3* deletion (Fig. 1*A*). This result is in accordance with the finding that *Tb*STT3A can utilize the $Man_5GlcNAc_2$ LLO substrates in *T. brucei* (19). Interestingly, in both the Δ*stt3*Δ*alg9* and the Δ*stt3*Δ*alg12* strains, all three *TbSTT3* paralogues complemented the deletion of endogenous OST activity (Fig. 1, *B* and *C*). In the Δ*stt3* strain, only *Tb*STT3B and *Tb*STT3C complement the *STT3* deletion, whereas *Tb*STT3A-expressing cells did not survive plasmid shuffling (Fig. 1*E*) (19). The inability of *Tb*STT3A to complement Δ*stt3* in the presence of $Glc_3Man_9GlcNAc_2$ LLOs was independent of the three terminal glucose residues on the LLO substrate because growth of Δ*stt3*Δ*alg6* cells expressing *Tb*STT3A was also not rescued (Fig. 1*D*). Our data confirm that the complementary activity of the different *T. brucei* STT3 proteins depended on the oligosaccharide structures of the substrate LLO.

Two distinct protein regions influence LLO specificity of TbSTT3B and TbSTT3C

*Tb*STT3B and *Tb*STT3C protein sequences are ~95% identical, and sequence differences cluster, in contrast to the *Tb*STT3A, in three distinct regions located in lumenal part of the proteins (Fig. 2). These two OSTs differ significantly with respect to their LLO substrate specificity: *Tb*STT3B does not accept $Man_5GlcNAc_2$ LLOs, in contrast to *Tb*STT3C. We therefore reasoned that one or a combination of these regions would account for different LLO specificities of *Tb*STT3B and *Tb*STT3C and that chimeric *Tb*STT3B/C proteins will be fold properly.

The differences in LLO specificity led to the inability of *Tb*STT3B to rescue the growth of the Δ*stt3*Δ*alg3* strain, whereas *Tb*STT3C was able to complement the *STT3* deletion of this strain. *Tb*STT3B-C chimeras were constructed by exchanging single regions of *Tb*STT3B by corresponding regions of *Tb*STT3C (*Tb*STT3B-1C, *Tb*STT3B-2C, and *Tb*STT3B-3C) or by combinations of two regions (*Tb*STT3B-1/2C, *Tb*STT3B-2/3C, and *Tb*STT3B-1/3C). To test the effect of the region exchange on LLO specificity, Δ*stt3*Δ*alg3* cells harboring the chimeric *Tb*STT3B-C constructs were subjected to 5-FOA-induced plasmid shuffling, and growth of the cells was

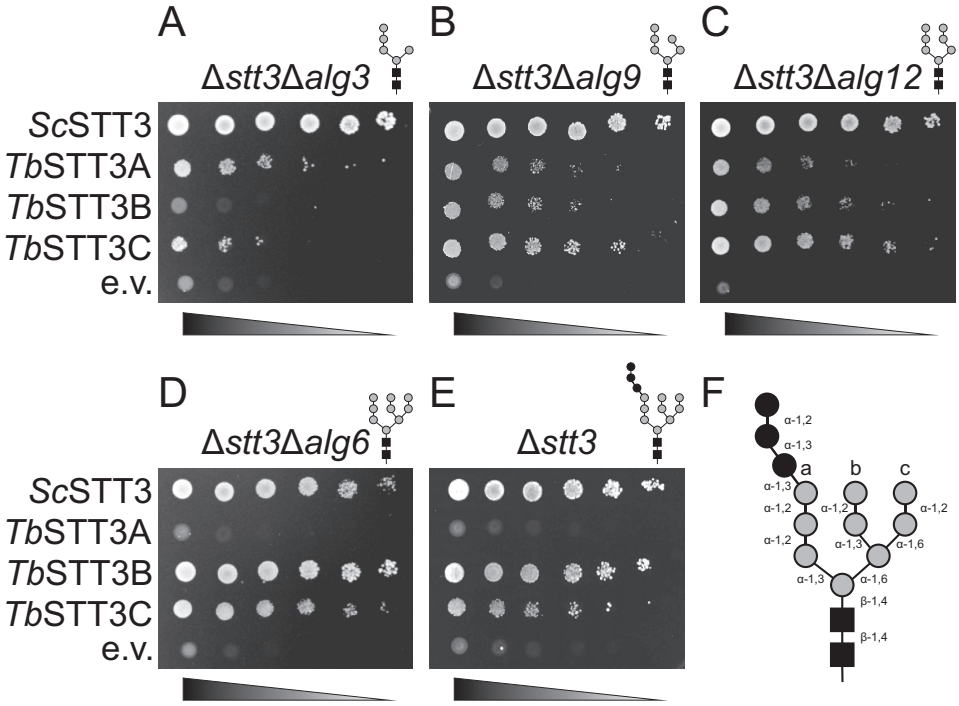


Figure 1. LLO specificity of *Tb*STT3 paralogues. Strains deleted for genomic *STT3* (Δ*stt3*) were complemented by a *URA3*-plasmid encoding yeast *STT3* (*Sc*STT3). The strains harbored a second *LEU2*-marked plasmid encoding *Tb*STT3A, *Tb*STT3B, *Tb*STT3C, *Sc*STT3, or no *STT3* gene (*e.v.*). Serial dilutions of Δ*stt3*Δ*alg3* (*A*), Δ*stt3*Δ*alg9* (*B*), Δ*stt3*Δ*alg12* (*C*), Δ*stt3*Δ*alg6* (*D*), and Δ*stt3* (*E*) were spotted onto 5-FOA-containing medium and incubated at 23 °C for different times to select for ura− cells. STT3s encoded on the *LEU2* plasmid complementing the genomic Δ*stt3* deletion support growth on 5-FOA-containing medium. LLO structures of the respective strains were depicted schematically. *F*, yeast LLO structure is represented by two GlcNAc (*black squares*), nine Man (*gray circles*), and three Glc (*black circles*) residues with the respective linkages indicated, and LLO branches are labeled by *a*–*c*.

assessed (Fig. 3*A*). The exchange of region 1 in *Tb*STT3B did not rescue the cell growth in the Δ*stt3*Δ*alg3* ($Man_5GlcNAc_2$), and the growth in the $Glc_3Man_9GlcNAc_2$ accumulating strain was reduced as well. In contrast, the exchange of region 2 (*Tb*STT3B-2C) allowed cell growth in both backgrounds comparable with *Tb*STT3C. Also, the exchange of region 3 (*Tb*STT3B-3C) supported growth of Δ*stt3*Δ*alg3* cells albeit at a somewhat reduced level. These observations were further substantiated by the *Tb*STT3B-C chimeras with two regions exchanged. Both chimeras containing region 2 (*Tb*STT3B-1/2C and *Tb*STT3B-2/3C) promoted growth similar to *Tb*STT3C, whereas the chimera with regions 1 and 3 (*Tb*STT3B-1/3C) resulted in a severe growth phenotype in Δ*stt3*Δ*alg3* cells, but not in cells with normal LLO biosynthesis (Fig. 3, *A* and *B*). We concluded that both regions 2 and 3 were required for optimal utilization of $Man_5GlcNAc_2$ LLO. However, region 2 seems to be more important in determining LLO substrate specificity of *Tb*STT3B and *Tb*STT3C describing an oligosaccharide recognition domain of this single subunit OST.

TbSTT3A, TbSTT3B, and TbSTT3C display differential preferences for polypeptide substrates

Having established that region two is responsible for providing specificity toward the LLO donor substrate, we also sought to identify the region(s) of the *Tb*STT3 paralogues that provide(s) specificity toward polypeptide substrates. In the first step, differences in the glycosylation efficiencies of the three *Tb*STT3 paralogues for different glycosylation sites of yeast proteins were addressed. It was previously reported that *Tb*STT3A and *Tb*STT3C modified more efficiently the glycosylation sites that are located in a sequence context with acidic amino acids, whereas *Tb*STT3B did not display a particular preference for specific amino acids in the local environment of the glycosylation site (19). A parallel reaction monitoring (PRM) mass spectrometry (MS)-based method (Fig. 4*A*) was used to determine the occupancy of glycosylation sites from different yeast membrane proteins, which represents the percentage of peptides modified with a glycan at a given glycosylation site compared with same peptide in wild-type cells. To monitor all three *T. brucei* OST proteins, we used the Δ*stt3*Δ*alg9* yeast strain to compare the glycosylation efficiencies of *Tb*STT3A, *Tb*STT3B, and *Tb*STT3C. This strain generates an LLO substrate compatible with all *T. brucei* OSTs, and the *N*-linked glycan is susceptible to EndoH digestion.

The stable isotope labeling with amino acids in cell culture (SILAC) coupled to the PRM MS-based technique (27) was used to analyze glycosylation occupancy at glycosylation sites on proteins in yeast microsomal membrane preparations. In short, a reference wild-type strain was grown in medium containing heavy isotope-labeled arginine and lysine, whereas the strains expressing *Tb*STT3s were grown in the corresponding medium using regular amino acids (light). Cells were mixed 1:1 and disrupted, and samples were enriched for the glycoprotein-rich membrane fraction. *N*-Linked glycans were cleaved by EndoH to maintain the first GlcNAc residue of the glycan on the protein. Proteins were digested enzymatically, and the resulting peptides were analyzed by liquid chromatography-electrospray-MS/MS. Corresponding light and heavy peptides were paired, and site occupancies relative to the reference

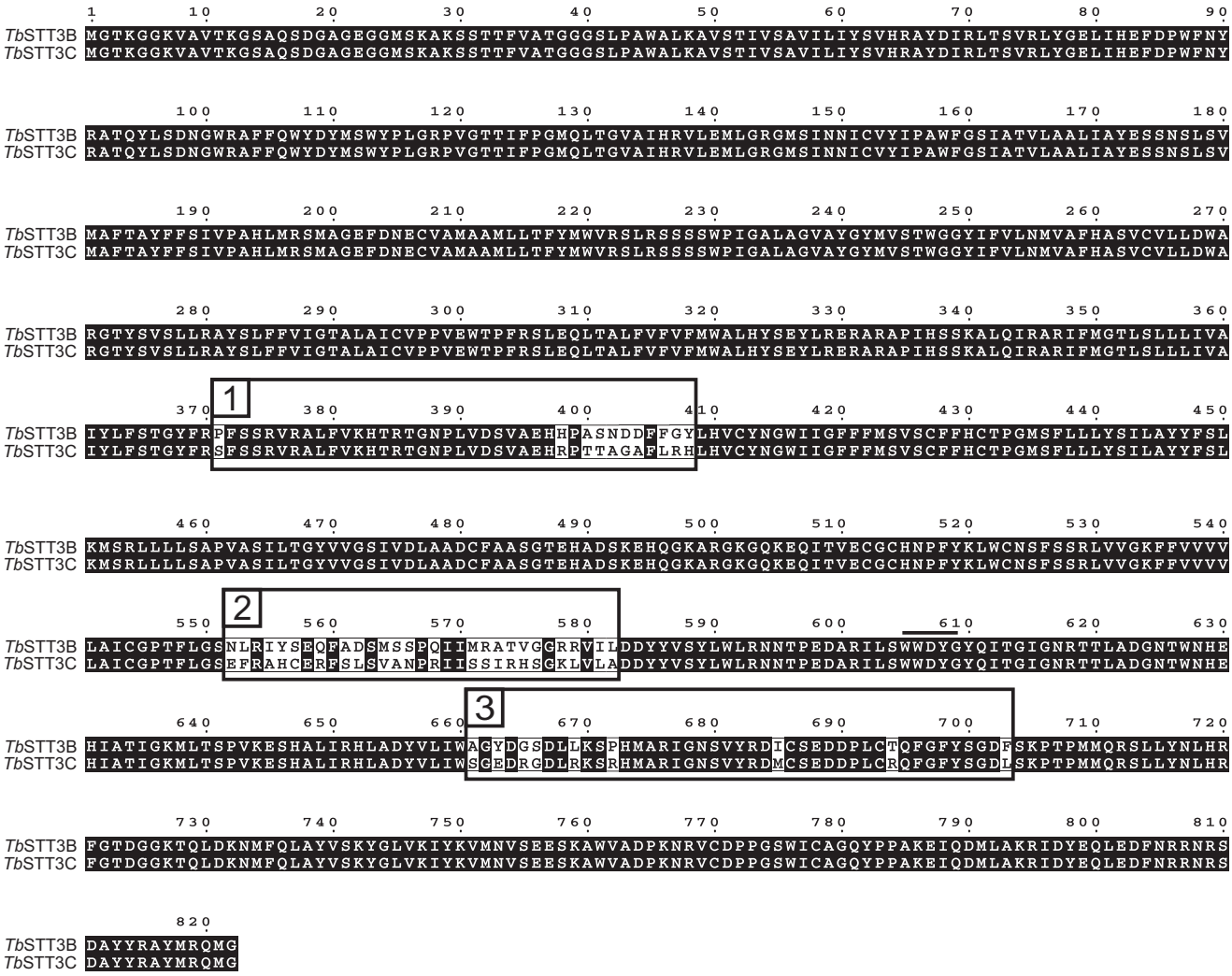


Figure 2. Protein sequence alignment of *Tb*STT3B and *Tb*STT3C. Protein sequences of *Tb*STT3B and *Tb*STT3C were aligned, and identical residues are displayed in *black*. Residues distinct in *Tb*STT3B and *Tb*STT3C are *labeled white*. The sequences of *Tb*STT3B and *Tb*STT3C show sequence differences in three discrete regions (*1–3*) labeled by a *rectangle*. All three regions are located in the lumenal part of the *Tb*STT3 proteins. The conserved WWD*X*G motif is marked by a *horizontal line* above the sequence.

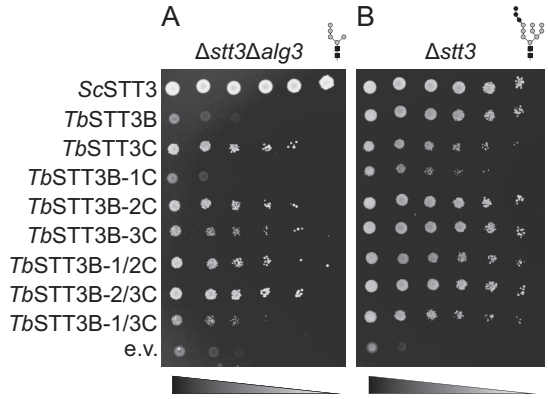


Figure 3. LLO specificity of *Tb*STT3s is provided by distinct regions in the proteins. Strains deleted for genomic STT3 (Δ*stt3*) were complemented by a *URA3* plasmid encoding yeast *STT3* (*Sc*STT3). The strains harbored a second *LEU2*-marked plasmid encoding *Tb*STT3B, *Tb*STT3C, and *Tb*STT3B-C chimeras (see text for description), *Sc*STT3, or no *STT3* gene (*e.v.*). Serial dilutions of Δ*stt3*Δ*alg3* (*A*) and Δ*stt3* (*B*) were spotted on 5-FOA containing medium and incubated at 23 °C for 7 days (Δ*stt3*) to 15 days (Δ*stt3*Δ*alg3*) to select for *Ura*$^-$ cells. STT3s encoded on the *LEU2*-plasmid complementing the genomic Δ*stt3* deletion support growth on 5-FOA-containing medium. LLO structures of the respective strains were depicted schematically.

strain were calculated for *Tb*STT3-expressing cells based on light/heavy (L/H) ratios of the peak area values (supplemental Table 1). Site occupancy reflected the preference of the OST for a given glycosylation site and its local environment. Glycosylation sites that are favored by a given OST will be glycosylated more efficiently (*i.e.* higher site occupancy) as compared with sites not located in a favored peptide sequence context.

Cluster analysis of the site occupancy data acquired for the *Tb*STT3 paralogues indicated that *Tb*STT3A and *Tb*STT3C glycosylation efficiency of the 55 analyzed sites were similar and the efficiency of glycosylation by *Tb*STT3B was distinct from *Tb*STT3A and *Tb*STT3C (Fig. 4*B*). This result confirmed previously reported differences in the glycosylation efficiency observed for the *Tb*STT3B and *Tb*STT3C (19).

We then examined the polypeptide substrate specificity of *Tb*STT3 paralogues in the context of sequence polarity. Sequence composition analysis was performed on glycosylation sequons themselves plus 10 residues downstream and upstream of the glycosylation sites. Amino acid residues were grouped based on their polarity into acidic (Asp and Glu), basic

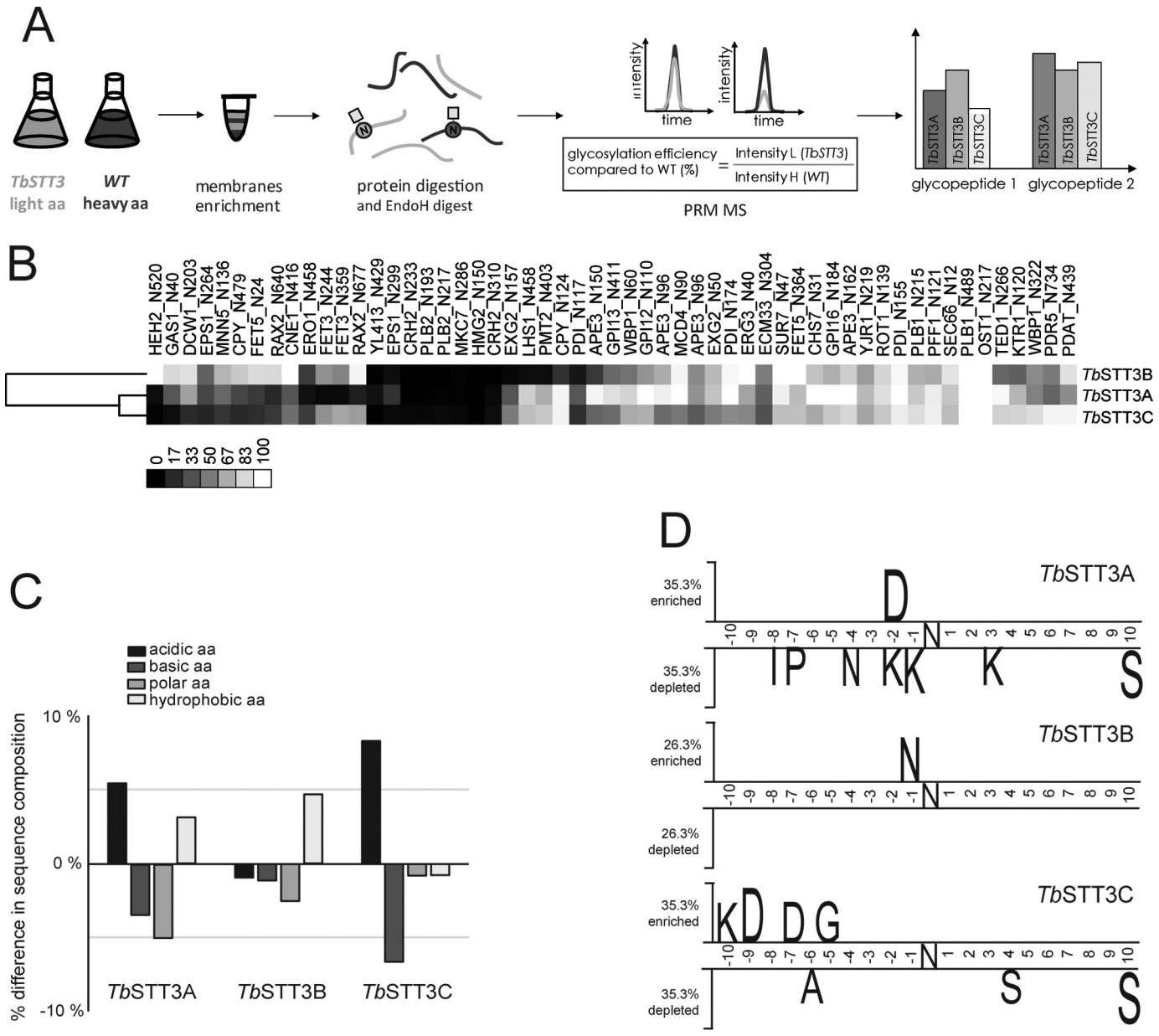


Figure 4. Glycosylation site occupancy and sequence composition analysis for *Tb*STT3s in Δ*stt3*Δ*alg9* cells. Δ*stt3*Δ*alg9* cells complemented with *Tb*STT3A, *Tb*STT3B, and *Tb*STT3C were grown in light medium and mixed 1:1 with the wild-type reference strain grown in heavy medium, and membrane-derived peptides were prepared. Peptide abundance was measured by PRM mass spectrometry. Intensity ratios of glycosylated light to heavy peptides were normalized for expression differences in *Tb*STT3-expressing cells and wild-type cells. The resulting ratios represent the site occupancy for the *Tb*STT3-expressing cells relative to the wild-type reference strain. Data are the mean of biological triplicates. *A,* schematic representation of SILAC PRM MS method. *B,* site occupancy values for *Tb*STT3A, *Tb*STT3B, and *Tb*STT3C are presented in a heatmap. Color is mapped from *black* (0%) to *white* (100%). Data are from supplemental Table 1. Data were used for cluster analysis. In cluster analysis samples with high similarity are close, and samples with low similarity are distant from each other. *C,* sequence composition analysis was performed where amino acids were grouped based on their polarity. Percentage change in respective ratio of each amino acid group between efficiently and poorly glycosylated sequences upstream of glycosylation sequons was calculated for each *Tb*STT3-expressing strain. Efficiently glycosylated sites are considered where glycosylation occupancy ratio compared with wild type is more than 0.75, and poorly glycosylated sites are considered where glycosylation occupancy ratio compared with wild type is less than 0.25. *D,* two-sample logo analysis (30) was used to calculate and visualize residues surrounding glycosylation sites that are significantly enriched in either efficiently or poorly glycosylated sites in each TbSTT3-expressing strain. Sequence size analyzed contained 10 amino acids upstream and downstream from the glycosylation site. Efficiently glycosylated sites are considered where glycosylation occupancy ratio compared with wild type is more than 0.75, and poorly glycosylated sites are considered where glycosylation occupancy ratio compared with wild type is less than 0.25.

(Arg, Lys, and His), polar (Ser, Thr, Asn, Glu, Cys, and Tyr), and hydrophobic (Ala, Val, Ile, Leu, Met, Phe, Trp, Pro, and Gly) groups. Sequences were divided into "efficiently" and "poorly" glycosylated when the glycosylation occupancy calculated was more than 85% or less than the 25% compared with reference strains, respectively. The ratio of each amino acid group was calculated compared with total number. Percentage change in respective ratios of each amino acid group between efficiently and poorly glycosylated sequences downstream and upstream of the glycosylation sequon was calculated for each *Tb*STT3-expressing strain. The analysis of sequences downstream of the glycosylation sites showed no apparent difference in sequence

specificity between different *Tb*STT3 paralogues (supplemental Table 2). All three enzymes showed preference for hydrophobic residues, whereas polar residues were not favored. The analysis of upstream sequences revealed that both *Tb*STT3A and *Tb*STT3C preferred acidic residues (Fig. 4*C*). Additionally, both paralogues showed disfavor toward basic residues. However, *Tb*STT3B showed no specific preference for any amino acid type. These results support a previously reported difference in the polypeptide acceptor substrate specificity, where *Tb*STT3A and *Tb*STT3C showed selectivity toward glycosylation sequons flanked by acidic residues, whereas *Tb*STT3B lacked any obvious preference (19). Furthermore, it has been confirmed that upstream amino acids play a dominant role in *Tb*STT3A acceptor peptide specificity (35).

Site occupancy data acquired for the *Tb*STT3 paralogues was further exploited to examine whether there is a particular preference for specific amino acids in the local environment of the glycosylation sites. To establish whether there are consensus sequences or patterns that control site-specific *N*-glycosylation, a two-sample logo analysis (30) was used to calculate and visualize statistically significant residues surrounding glycosylation sites (Fig. 4*D*). In agreement with sequence composition analysis, both *Tb*STT3A and *Tb*STT3C showed efficient glycosylation of sites containing aspartic acid. Sequences enriched in aspartic acid at the −2 position are more likely to be efficiently glycosylated by *Tb*STT3A, whereas *Tb*STT3C preferred sequences enriched in aspartic acid at the −9 position. Although the effect of distal residues on the glycosylation occupancy (*e.g.* aspartic acid at −9 position) needs to be confirmed by mutagenesis studies, the preference for an acidic residue at the −2 position has been demonstrated *in vitro* (31). Similar to previous data, *Tb*STT3B showed no specific preference. Preference for *Tb*STT3A for amino acids immediately adjacent to the glycosylation site is supported by an extensive analysis of native and artificial polypeptide substrates in *T. brucei* cells (35).

Distinct protein region influences polypeptide specificity of TbSTT3B and TbSTT3C

To define the polypeptide substrate specificity more closely, we used *Tb*STT3B-C chimeras to identify regions of the OST proteins that influenced the preference for certain glycosylation sites. Because of the fact that the chimera *Tb*STT3B-1C yielded poor growth of the strains, we took advantage of the stabilizing property of the concomitant exchange of region 3 and compared the *Tb*STT3B-1/3C (= *Tb*STT3C-2B) with the *Tb*STT3B-2/3C (= *Tb*STT3C-1B) chimera that yielded similar growth in Δ*stt3*Δ*alg9* cells (supplemental Fig. 1). Glycosylation efficiencies of different chimeras were determined using the MS-based method described above, and the cluster analysis of the site occupancy data was performed (supplemental Table 3). *Tb*STT3B-1/3C displayed the highest similarity with *Tb*STT3C, whereas the corresponding chimera *Tb*STT3B-2/3C showed only little similarity with *Tb*STT3C (Fig. 5*A*). Exchange of region 1 seems to influence glycosylation efficiency of the two TbSTT3 paralogues examined.

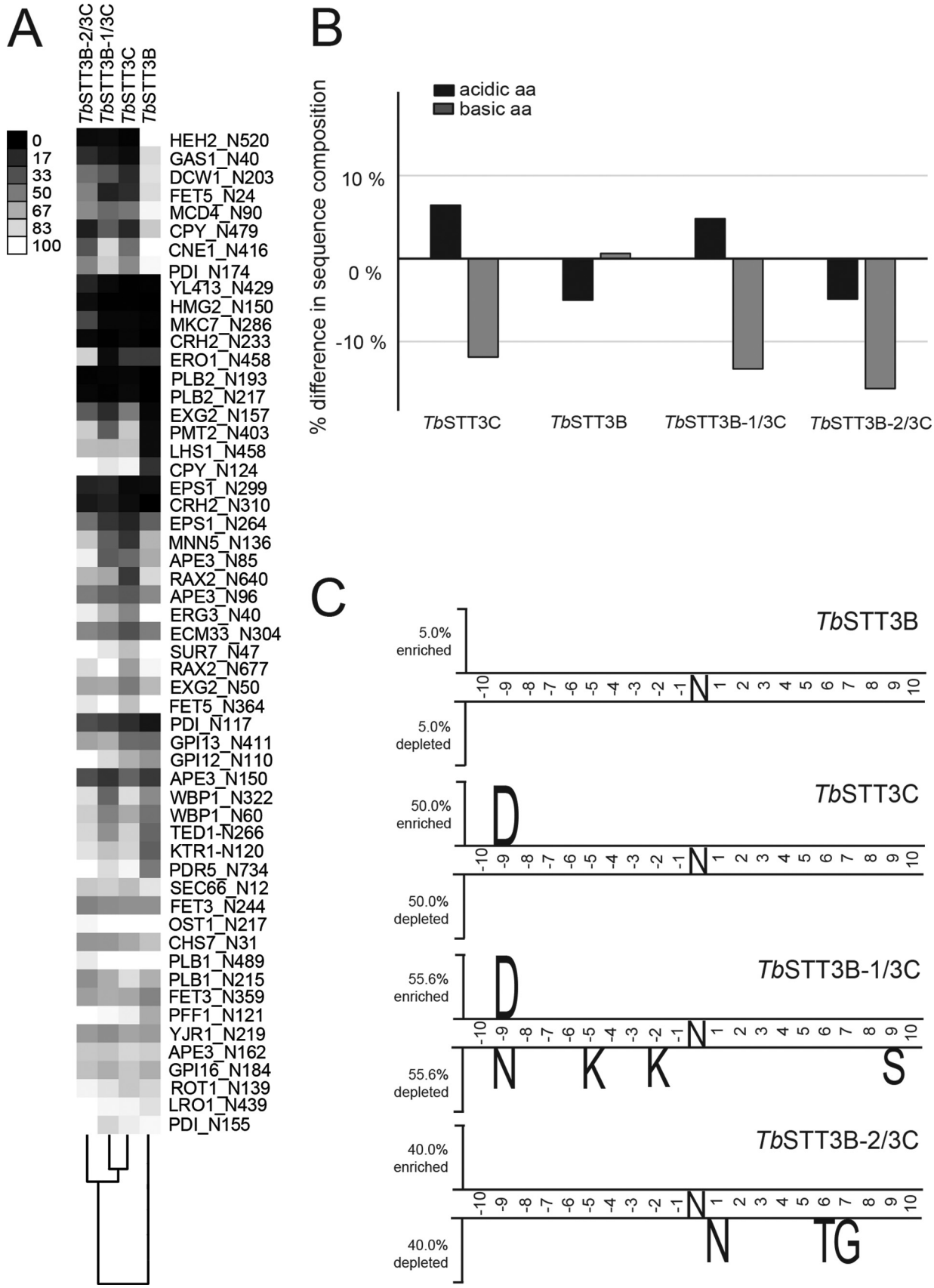
We focused the analysis only on the sites that were differentially glycosylated by *Tb*STT3B and *Tb*STT3C. Sequence composition analysis was performed where the percentage change in respective ratios of acidic and basic amino acid groups between efficiently and poorly glycosylated sequences upstream of the glycosylation sequon was calculated for each *Tb*STT3-expressing strain. As expected, *Tb*STT3C showed preference for the acidic residues, whereas in this analysis, these were disfavored by *Tb*STT3B enzyme (Fig. 5*B*). Similar to *Tb*STT3C, the *Tb*STT3B-1/3C chimera displayed preference for acidic sequences upstream of the glycosylation sites. Interestingly, a two-sample logo analysis revealed that both the *Tb*STT3C paralogue and *Tb*STT3B-1/3C chimera showed preference for an aspartic acid at the same position upstream of the glycosylation sequon, although, as mentioned before, the effect of distal residues on the glycosylation occupancy needs to be further investigated (Fig. 5*C*). Similar to previous data, *Tb*STT3B showed no specific preference, and the same was true for *Tb*STT3B-2/3C chimera. Involvement of region 1 in sequon specificity of *Tb*STT3B and *Tb*STT3C has been previously demonstrated, whereupon genetic rearrangements a chimeric gene was generated containing the first variable region of *Tb*STT3C flanked by *Tb*STT3B sequences. The chimeric *Tb*STT3B/C/B protein described in this study showed much less efficient recognition of the native substrate of *Tb*STT3B, whereas it appeared to have attained a peptide acceptor specificity more similar to *Tb*STT3A than *Tb*STT3B (21).

Discussion

Functional expression of single subunit OSTs from kinetoplastids in *S. cerevisiae* has proven to be a useful model system to study the properties of STT3s from *L. major*, *T. cruzi*, and *T. brucei* (16–19). Yeast genetics methods allow manipulations of the LLO biosynthesis pathway to generate specific intermediate oligosaccharide structures (26) that can be used to study the influence of altered LLO substrates on OSTs. Consequences of such altered substrates for *N*-glycosylation can be monitored by analyzing *N*-glycoproteins, the products of the OST-catalyzed reaction. Both single *N*-glycoproteins like carboxypeptidase Y (32, 33) and MS-based methods, which allow a broader view on many *N*-glycoproteins at the same time, were used to study consequences of alterations in the *N*-glycosylation process (27). The combination of genetics and analytical tools available in *S. cerevisiae* thus represents an excellent system to perform reverse genetics approaches to study particular OST features.

In vivo analysis of the *T. brucei* VSG221 protein showed that *Tb*STT3A transfers $Man_5GlcNAc_2$ glycans to protein, whereas *Tb*STT3B prefers $Man_9GlcNAc_2$ as a substrate for *N*-glycosylation (19). Analysis of VSG221 proteins in *T. brucei* *Tb*ALG3$^{-/-}$ and *Tb*ALG12$^{-/-}$ mutant strains revealed also that LLO intermediates can serve as substrates for both *Tb*STT3A and *Tb*STT3B, although with reduced efficiency. It was hypothesized that efficient glycosylation of VSG221 by *Tb*STT3A correlates with the absence of the LLO c-branch, whereas for *Tb*STT3B the presence of the c-branch is an important determinant to improve glycosylation (19, 36).

Analysis of the LLO specificities of *Tb*STT3A and *Tb*STT3B in *T. brucei* was possible, because the two VSG221 glycosylation sites get selectively modified either by *Tb*STT3A with $Man_5GlcNAc_2$ glycans or by *Tb*STT3B with $Man_9GlcNAc_2$



oligosaccharides. This selectivity for a specific glycosylation sequon is provided by the distinct polypeptide specificities of *Tb*STT3A and *Tb*STT3B (19). Because *Tb*STT3C was not expressed in *T. brucei*, its substrate specificity was investigated in the heterologous yeast expression system. *Tb*STT3C shared the preference for acidic sequons with *Tb*STT3A but used $Man_9GlcNAc_2$ as LLO substrate, as observed with *Tb*STT3B (19).

Our detailed investigations on LLO specificity in the yeast *in vivo* system confirmed the results obtained in *T. brucei* for *Tb*STT3A and *Tb*STT3B. Furthermore, we demonstrated that the inability of *Tb*STT3A to complement the *STT3* deletion of yeast cells was independent of LLO glucosylation. Because the terminal Glc residues of the LLO a-branch did not influence *Tb*STT3A, although *T. brucei* synthesizes only non-glucosylated LLOs (22), an interaction of *Tb*STT3A with the terminal Man residue of the LLO a-branch seems unlikely. The inability of *Tb*STT3A to support growth of the Δ*stt3* and Δ*stt3*Δ*alg6* strains was rather due to the presence of LLO c-branch mannoses, which seemed to prevent efficient glycosylation by *Tb*STT3A. *Tb*STT3B was able to support growth of all strains tested with the exception of Δ*stt3*Δ*alg3*. In *T. brucei*, *Tb*STT3B modified VSG221 in a *Tb*ALG3$^{-/-}$ strain, although with reduced efficiency (36). The inability of *Tb*STT3B to rescue the growth of Δ*stt3*Δ*alg3* yeast cells was therefore likely to result from reduced overall glycosylation levels, which were too low to allow survival of the yeast cells, rather than the inability to use $Man_5GlcNAc_2$ LLOs as substrate in the heterologous host. Opposed to *Tb*STT3A and *Tb*STT3B, both of which displayed distinct LLO donor substrate preferences, all genetically tailored LLOs served as a substrate for *Tb*STT3C. Although *Tb*STT3 paralogues could utilize a range of LLO substrates to glycosylate proteins, no intermediate glycan structures are transferred to protein by *Tb*STT3A and *Tb*STT3B unless *ALG* mutations were introduced (19, 25, 36). This indicates that additional factors like availability or accessibility of LLO biosynthesis intermediate and k_m values of the *ALG* enzymes fine tune the glycosylation machinery leading to the specific transfer of $Man_5GlcNAc_2$ LLOs by *Tb*STT3A and $Man_9GlcNAc_2$ oligosaccharides by *Tb*STT3B (19).

The heterologous expression system allowed us to perform targeted structure–function analyses. Region 2, which was found to be important for LLO specificity, is located in the C-terminal part of the *Tb*STT3 protein after the last predicted transmembrane helix. This C-terminal domain was predicted to be localized in the ER lumen due to the presence of the highly conserved WWD*X*G motif that is important for the glycosylation reaction of OSTs (5, 37, 38). The results presented here identify for the first time regions of the *Tb*STT3 proteins that are important for LLO specificity and provide a basis to further investigate the molecular mechanisms and contribution of single amino acids in OST LLO interaction.

Glycosylation efficiency of a given OST can be determined by analyzing site occupancy for different substrate polypeptides (17, 19, 39–41). Here, we compared the site occupancies for the three *Tb*STT3 paralogues in the Δ*stt3*Δ*alg9* strain relative to a wild-type reference strain. Our results confirmed previous observations, made in two different expression hosts, that *Tb*STT3A and *Tb*STT3C have similar polypeptide substrate preferences (19). We further found that region 1 influences the glycosylation efficiency of certain polypeptide substrate glycosylation sites. *PglB*, the bacterial homologue of *STT3*, interacts with the threonine/serine residue of the glycosylation sequon of its peptide substrate via residues of the conserved WWD*X*G motif, which determines the specificity for the sequon sequence. A periplasmic loop connecting transmembrane helices 9 and 10 (*i.e.* external loop 5; EL5) also showed significant interaction with the peptide substrate. The C-terminal part of EL5 pins the peptide against the periplasmic domain, but it also contains the conserved residue Glu-319 that is part of the catalytic site of PglB (38). In the *Tb*STT3 paralogues, the region surrounding the WWD*X*G motif is completely conserved. Therefore, it seems unlikely that this region is responsible for the differences observed in site occupancy between *Tb*STT3B and *Tb*STT3C. Sequence alignments between *PglB* from *Campylobacter lari* and *Tb*STT3B and *Tb*STT3C showed that the conserved residue Glu-319 of *PglB* has an equivalent glutamic acid residue in the *Tb*STT3s that is located in region 1 of *Tb*STT3B and *Tb*STT3C (*i.e.* Glu-396). Hence, it is tempting to speculate that region 1 could be the functional equivalent to the EL5 described in *PglB*. Region 1 might interact with polypeptide substrates and modulate the glycosylation efficiency of *Tb*STT3s. Interestingly, toward the C-terminal end of region one, both *Tb*STT3A and *Tb*STT3C have conserved sequences, although the amino acid residues of *Tb*STT3B at these positions are different. This seemed to coincide with the observed differences in site occupancy and sequence composition for *Tb*STT3A and *Tb*STT3C compared with *Tb*STT3B. We showed that amino acids upstream of the glycosylation site play a dominant role in polypeptide specificity of *Tb*STT3. Furthermore, we were able to depict specific amino acid residues in polypeptide substrate sequences that increase the chances of such substrates being efficiently glycosylated by different *Tb*STT3 paralogues. Sequences enriched in aspartic acid at the −2 position are more likely to be efficiently glycosylated by *Tb*STT3A, whereas *Tb*STT3C preferred sequences enriched for aspartic acid at the −9 position. These conclusions are supported by the analysis of *Tb*STT3A and *Tb*STT3B specificity in

Figure 5. Glycosylation site occupancy and sequence composition analysis for *Tb*STT3B-1/3C and *Tb*STT3B-2/3C chimeras in Δ*stt3*Δ*alg9* cells. Δ*stt3*Δ*alg9* cells complemented with *Tb*STT3B, *Tb*STT3C, *Tb*STT3B-1/3C, and *Tb*STT3B-2/3C were grown in light medium and mixed 1:1 with the wild-type reference strain grown in heavy medium, and membrane-derived peptides were prepared. Intensity ratios of glycosylated light to heavy peptides were normalized for expression differences in *Tb*STT3-expressing cells and wild-type cells. The resulting ratios represent the site occupancy for the *Tb*STT3-expressing cells relative to the wild-type reference strain (reported in %). *A,* site occupancy values for *Tb*STT3B, *Tb*STT3C, *Tb*STT3B-1/3C, and *Tb*STT3B-2/3C were used for cluster analysis. In cluster analysis samples with high similarity are close whereas samples with low similarity are distant from each other. *B,* sequence composition analysis was performed where amino acids were grouped based on their polarity. Percentage change in respective ratio of each amino acid group between efficiently and poorly glycosylated sequences upstream of glycosylation sequon was calculated for each *Tb*STT3-expressing strain. *C,* two-sample logo analysis (30) was used to visualize differences between efficiently and poorly glycosylated sequences surrounding the glycosylation sites for each TbSTT3-expressing strain.

T. brucei cells directly (35). Based on a larger data set, a preference for acidic acceptor sequences of *Tb*STT3A was observed. Molecular modeling let these authors conclude that sequences in region 1 are responsible for this substrate specificity. Our experimental data support this conclusion. In addition, *in vitro* analysis of *Tb*STT3A substrate specificity revealed an activating effect of an acidic residue at position −2 (31). More experimental work and crystal structures of eukaryotic single subunit OSTs will provide insights into the molecular basis of this polypeptide substrate specificity.

Experimental procedures

Media, yeast strains, and plasmids

Standard yeast media and molecular biology methods were used (42). For maintenance of plasmids, cells were grown in appropriate synthetic medium lacking amino acids necessary for selection. To generate haploid double mutant strains with genomic deletions of *STT3* and *ALG3, ALG9,* and *ALG6* genes to test the complementation by different TbSTT3s, haploid mutant strains with deletions in *ALG3, ALG6,* and *ALG9* genes were purchased (Euroscarf): Y13108 (*alg3*Δ*::KanMX4*), Y01993 (*alg9*Δ*::KanMX4*), and Y11778 (*alg6*Δ*::KanMX4*). These strains were individually mated with YBS10 and YBS11 strains, respectively (17); sporulation was induced, and tetrads were dissected on YPD plates containing 1 M sorbitol. Haploid spores harboring the *STT3* plasmid and the two deletions in the *STT3* and *ALG3, ALG6,* or *ALG9* loci were identified on G418-containing media in non-parental ditype tetrads. The absence of the *ALG3, ALG6,* and *ALG9* genes in the double mutant strains was confirmed by PCR. The Δ*stt3*Δ*alg12* strain was generated from YG2082 by exchanging LmSTT3D by the plasmid encoding the yeast STT3 locus (17, 43). To test complementation of the *STT3* deletion by *Tb*STT3 paralogues, the following strains were used: Δ*stt3* (YBS11) (*STT3::kanMX4 MAT*a *his3*Δ*1 leu2*Δ*0 met15*Δ*0 ura3*Δ*0 lys2*Δ*0*) (17); Δ*stt3*Δ*alg6* (*stt3*Δ*::kanMX4 alg6*Δ*::kanMX4 MAT*a *his3*Δ*1 leu2*Δ*0 met15*Δ*0 ura3*Δ*0 lys2*Δ*0*); Δ*stt3*Δ*alg12 (stt3*Δ*::kanMX4 alg12*Δ*::kanMX4 MATa his3*Δ*1 leu2*Δ*0 met15*Δ*0 ura3*Δ*0 lys2*Δ*0*); Δ*stt3*Δ*alg9* (*stt3*Δ*::kanMX4 alg9*Δ*::kanMX4 MAT*α *his3*Δ*1 leu2*Δ*0 met15*Δ*0 ura3*Δ*0 lys2*Δ*0*); Δ*stt3*Δ*alg3* (*stt3*Δ*::kanMX4 alg3*Δ*::kanMX4 MAT*a *his3*Δ*1 leu2*Δ*0 ura3*Δ*0 lys2*Δ*0*). The *STT3* deletion was complemented by a plasmid encoding the yeast *STT3* locus in a URA3-marked YEp352 vector (43). These cells were transformed with *T. brucei* STT3 paralogues encoded by LEU2-marked plasmids under the control of the yeast GPD promoter (19). The yeast STT3 locus was cloned as a HindIII fragment from the yeast *STT3* locus harboring plasmid (43) into the pRS425GPD vector as control for the plasmid shuffling procedure. Cells were grown after plasmid shuffling in appropriate defined medium containing 1 M sorbitol. *Tb*STT3B-C chimera plasmids were generated by homologous recombination. PCR fragments encoding the different regions (regions 1–3) were generated with the following primers using either *Tb*STT3B or *Tb*STT3C plasmids as template: 5′-AGCGCAACTACAGAGAAC-3′; 5′-ACAGATGGCAAGGACAAC-3′; 5′-CGTTCCTGCTGTTGTACTC-3′; 5′-GCTATGTGCTCGTGATTCC-3′; 5′-CTGAAGATGCCCGTATTCTC-3′; and 5′-CGCACGATAATAAGCGTCAC-3′. PCR fragments were designed to have overlapping sequences with the vector or a PCR fragment containing a neighboring region. To assemble the different chimeras, *Tb*STT3B vector was gapped with NcoI and BglII, and yeast cells were co-transformed with the gapped vector and all possible combinations of PCR fragments encoding for the three different regions from either *Tb*STT3B or *Tb*STT3C for homologous recombination. Plasmids were isolated from yeast and amplified in *E. coli* DH5α, and correct chimera assembly was confirmed by sequencing the inserted DNA. For overexpression of *ALG12,* previously described constructs were used (28). Analysis of glycosylation site occupancy was performed with three biological replicates of the Δ*stt3*Δ*alg9* strain harboring *Tb*STT3A, *Tb*STT3B, *Tb*STT3C, or *Tb*STT3B-C chimera plasmids. KP4 (MATα *his3*Δ*1 leu2*Δ*0 lys2*Δ*0 ura3*Δ*0 arg4*Δ*::0*) was used as reference strain.

Specificity for lipid-linked oligosaccharides

Lipid-linked oligosaccharide specificity was tested by plasmid shuffling with the respective Δ*stt3*Δ*alg*-double mutant strains harboring both the *URA3* marked p*Sc*STT3 plasmid and the *LEU2* marked *Tb*STT3 or *Sc*STT3 encoding plasmids. These cells were subjected as 1:10 serial dilutions of 10^6 cells/ml to minimal medium containing 1 mg/ml 5-FOA (44) and 1 M sorbitol. The presence of 5-FOA allowed the selection of cells that lost the Ura^+ p*Sc*STT3 plasmid. These Ura^- cells only survived when the *STT3* genes encoded on the *LEU2*-plasmid could complement the yeast *STT3* deletion. Strains were incubated at 23 °C for 7–15 days depending on cell growth.

Sample preparation for mass spectrometry

Cells were grown at 25 °C in synthetic complete medium (0.67% (w/v) yeast nitrogen base, 2% (w/v) glucose with appropriate amino acid supplements) containing either 20 mg/liter light L-[$^{12}C_6$,$^{14}N_2$]lysine and L-[$^{12}C_6$]arginine or heavy L-[$^{13}C_6$,$^{15}N_2$]lysine and L-[$^{13}C_6$]arginine (Cambridge Isotope Laboratories). Cells were harvested in early log phase ($A_{600\,nm}$ = 0.8–1.2) by centrifugation and mixed 1:1 (w/w) for membrane protein preparation. Membrane proteins were prepared as described (45). Shortly, cells were lysed, and microsomal fractions were collected by high-spin centrifugation at 16,000 × *g*, 4 °C for 20 min. Proteins were processed using the filter-assisted sample preparation protocol (46). After reduction and alkylation, proteins were digested with Lys-C (20 μg/ml; Wako Pure Chemical, Richmond, VA) and trypsin (20 μg/ml; Promega) endopeptidases. Protein digestion was directly followed by EndoH treatment (500 units; New England Biolabs). Peptides were desalted using C18 ZipTips (Millipore) and dried using speed vacuum. Desalted peptides were resuspended in ACN/H_2O (3:97 (v/v)) with formic acid (0.1% (v/v)) and analyzed by LC-ESI-MS/MS.

Mass spectrometry analysis

MS analysis was performed by LC-ESI-MS/MS in PRM mode using a Q Exactive HF instrument (Thermo Fisher Scientific) coupled to ACQUITY UPLC system (Waters). Peptides were separated on HSS T3 column (78 μm × 150 mm, 1.8 μm) packed with C18 material (Waters). Peptides were eluted using

the gradient of 2–35% solvent B (99% (v/v) ACN, 0.1% (v/v) formic acid) over 90 min at a flow rate of 0.3 μl/min. All samples were analyzed using two PRM methods at an Orbitrap resolution of 30,000 or 60,000, based on scheduled inclusion lists containing the 175 and 128 target precursor ions, respectively, including retention time iRT standard peptides (Biognosys) (supplemental Table 4). The full scan event was collected using *m/z* 50–1400 mass selection, an Orbitrap resolution of 60,000 (at *m/z* 400), target automatic gain control (AGC) value of 3×10^6, and a maximum injection time of 30 ms. The PRM scan events used an Orbitrap resolution of 30,000 or 60,000, maximum fill time of 30 or 110 ms, respectively, an AGC value of 1×10^6 and with an isolation width of 2 *m/z*. Fragmentation was performed with a normalized collision energy of 28, and MS/MS scans were acquired with a starting mass of *m/z* 150. Scan windows were set to 10 min for each peptide in the final PRM method to ensure the measurement of 6–10 points per LC peak per transition.

Data processing and analysis

Skyline software (version 2.6.0) with standard settings was used for data processing (47). Briefly, raw MS data files were imported, and the peaks were manually inspected and adjusted to ensure proper peak picking and peak integration. The resulting light to heavy intensity ratio (L/H) for glycopeptides modified with HexNAc was used to calculate the relative site occupancy for the given peptide/glycosylation site. The relative site occupancy was normalized for expression differences between the heavy (H)-labeled reference wild-type strain and the *Tb*STT3-expressing light (L) strains by dividing the L/H intensity ratio for the occupied glycopeptide by the median of L/H intensity ratios reported for all non-glycopeptides (*i.e.* not containing an N*X*(T/S) sequon) from the same protein as the glycopeptide. Cluster analysis was performed with Cluster 3.0 software (library version 1.50) (29, 34) using the Spearman rank correlation to calculate similarity between site occupancy data for different *Tb*STT3s. Hierarchical clustering with the single linkage method was used to generate a dendrogram visualized with Java TreeView 1.1.6 software. Statistical tests used to analyze significant differences indicated in the respective figure legends were performed using *t* test in Microsoft Excel.

Author contributions—K. P., J. B., and M. A. conceived and coordinated the study and wrote the paper. R. G. assisted in conceiving the study. G. R. and M. P. contributed to growth assay analysis. All authors reviewed the results and approved the final version of the manuscript.

Acknowledgments—We thank N. Selevsek and the Functional Genomics Center Zurich for help with mass spectrometry.

References

1. Lowe, J. B., and Marth, J. D. (2003) A genetic approach to mammalian glycan function. *Annu. Rev. Biochem.* **72,** 643–691
2. Helenius, A., and Aebi, M. (2004) Roles of *N*-linked glycans in the endoplasmic reticulum. *Annu. Rev. Biochem.* **73,** 1019–1049
3. Burda, P., and Aebi, M. (1999) The dolichol pathway of *N*-linked glycosylation. *Biochim. Biophys. Acta* **1426,** 239–257
4. Gavel, Y., and von Heijne, G. (1990) Sequence differences between glycosylated and non-glycosylated Asn-*X*-Thr/Ser acceptor sites: implications for protein engineering. *Protein Eng.* **3,** 433–442
5. Yan, Q., and Lennarz, W. J. (2002) Studies on the function of oligosaccharyl transferase subunits. Stt3p is directly involved in the glycosylation process. *J. Biol. Chem.* **277,** 47692–47700
6. Kelleher, D. J., Karaoglu, D., Mandon, E. C., and Gilmore, R. (2003) Oligosaccharyltransferase isoforms that contain different catalytic STT3 subunits have distinct enzymatic properties. *Mol. Cell* **12,** 101–111
7. Nilsson, I., Kelleher, D. J., Miao, Y., Shao, Y., Kreibich, G., Gilmore, R., von Heijne, G., and Johnson, A. E. (2003) Photocross-linking of nascent chains to the STT3 subunit of the oligosaccharyltransferase complex. *J. Cell Biol.* **161,** 715–725
8. Pathak, R., Hendrickson, T. L., and Imperiali, B. (1995) Sulfhydryl modification of the yeast Wbp1p inhibits oligosaccharyl transferase activity. *Biochemistry* **34,** 4179–4185
9. Spirig, U., Bodmer, D., Wacker, M., Burda, P., and Aebi, M. (2005) The 3.4-kDa Ost4 protein is required for the assembly of two distinct oligosaccharyltransferase complexes in yeast. *Glycobiology* **15,** 1396–1406
10. Wilson, C. M., Roebuck, Q., and High, S. (2008) Ribophorin I regulates substrate delivery to the oligosaccharyltransferase core. *Proc. Natl. Acad. Sci. U.S.A.* **105,** 9534–9539
11. Schulz, B. L., Stirnimann, C. U., Grimshaw J. P., Brozzo, M. S., Fritsch, F., Mohorko, E., Capitani, G., Glockshuber, R., Grütter, M. G., and Aebi, M. (2009) Oxidoreductase activity of oligosaccharyltransferase subunits Ost3p and Ost6p defines site-specific glycosylation efficiency. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 11061–11066
12. Roboti, P., and High, S. (2012) Keratinocyte-associated protein 2 is a *bona fide* subunit of the mammalian oligosaccharyltransferase. *J. Cell Sci.* **125,** 220–232
13. Kelleher, D. J., and Gilmore, R. (2006) An evolving view of the eukaryotic oligosaccharyltransferase. *Glycobiology* **16,** 47R–62R
14. Wacker, M., Linton, D., Hitchen, P. G., Nita-Lazar, M., Haslam, S. M., North, S. J., Panico, M., Morris, H. R., Dell, A., Wren, B. W., and Aebi, M. (2002) *N*-Linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science* **298,** 1790–1793
15. Calo, D., Kaminski, L., and Eichler, J. (2010) Protein glycosylation in Archaea: sweet and extreme. *Glycobiology* **20,** 1065–1076
16. Castro, O., Movsichoff, F., and Parodi, A. J. (2006) Preferential transfer of the complete glycan is determined by the oligosaccharyltransferase complex and not by the catalytic subunit. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 14756–14760
17. Nasab, F. P., Schulz, B. L., Gamarro, F., Parodi, A. J., and Aebi, M. (2008) All in one: *Leishmania major* STT3 proteins substitute for the whole oligosaccharyltransferase complex in *Saccharomyces cerevisiae*. *Mol. Biol. Cell.* **19,** 3758–3768
18. Hese, K., Otto, C., Routier, F. H., and Lehle, L. (2009) The yeast oligosaccharyltransferase complex can be replaced by STT3 from *Leishmania major*. *Glycobiology* **19,** 160–171
19. Izquierdo, L., Schulz, B. L., Rodrigues, J. A., Güther, M. L., Procter, J. B., Barton, G. J., Aebi, M., and Ferguson, M. A. (2009) Distinct donor and acceptor specificities of *Trypanosoma brucei* oligosaccharyltransferases. *EMBO J.* **28,** 2650–2661
20. Castillo-Acosta, V. M., Ruiz-Pérez, L. M., Etxebarria, J., Reichardt, N. C., Navarro, M., Igarashi, Y., Liekens, S., Balzarini, J., and González-Pacanowska, D. (2016) Carbohydrate-binding non-peptidic pradimicins for the treatment of acute sleeping sickness in murine models. *PLoS Pathog.* **12,** e1005851
21. Castillo-Acosta, V. M., Vidal, A. E., Ruiz-Pérez, L. M., Van Damme, E. J., Igarashi, Y., Balzarini, J., and González-Pacanowska, D. (2013) Carbohydrate-binding agents act as potent trypanocidals that elicit modifications in VSG glycosylation and reduced virulence in *Trypanosoma brucei*. *Mol. Microbiol.* **90,** 665–679
22. Samuelson, J., Banerjee, S., Magnelli, P., Cui, J., Kelleher, D. J., Gilmore, R., and Robbins, P. W. (2005) The diversity of dolichol-linked precursors to Asn-linked glycans likely results from secondary loss of sets of glycosyltransferases. *Proc. Natl. Acad. Sci. U.S.A.* **102,** 1548–1553

23. Parodi, A. J., Quesada Allue, L. A., and Cazzulo, J. J. (1981) Pathway of protein glycosylation in the trypanosomatid *Crithidia fasciculata*. *Proc. Natl. Acad. Sci. U.S.A.* **78,** 6201–6205
24. de la Canal, L., and Parodi, A. J. (1987) Synthesis of dolichol derivatives in trypanosomatids. Characterization of enzymatic patterns. *J. Biol. Chem.* **262,** 11128–11133
25. Izquierdo, L., Mehlert, A., and Ferguson, M. A. (2012) The lipid-linked oligosaccharide donor specificities of *Trypanosoma brucei* oligosaccharyltransferases. *Glycobiology* **22,** 696–703
26. Jakob, C. A., Burda, P., te Heesen, S., Aebi, M., and Roth, J. (1998) Genetic tailoring of *N*-linked oligosaccharides: the role of glucose residues in glycoprotein processing of *Saccharomyces cerevisiae in vivo*. *Glycobiology* **8,** 155–164
27. Poljak, K., Selevsek, N., Ngwa, E. M., Grossmann, J., Losfeld, M. E., and Aebi, M. (2017) Quantitative profiling of *N*-linked glycosylation machinery in yeast *Saccharomyces cerevisiae*. *Mol. Cell. Proteomics* RA117.000096
28. Burda, P., Jakob, C. A., Beinhauer, J., Hegemann, J. H., and Aebi, M. (1999) Ordered assembly of the asymmetrically branched lipid-linked oligosaccharide in the endoplasmic reticulum is ensured by the substrate specificity of the individual glycosyltransferases. *Glycobiology* **9,** 617–625
29. Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U.S.A.* **95,** 14863–14868
30. Vacic, V., Iakoucheva, L. M., and Radivojac, P. (2006) Two sample logo: a graphical representation of the differences between two sets of sequence alignments. *Bioinformatics* **22,** 1536–1537
31. Ramírez, A. S., Boilevin, J., Biswas, R., Gan, B. H., Janser, D., Aebi, M., Darbre, T., Reymond, J.-L., and Locher, K. P. (2017) Characterization of the single-subunit oligosaccharyltransferase STT3A from *Trypanosoma brucei* using synthetic peptides and lipid-linked oligosaccharide analogs. *Glycobiology* **27,** 525–535
32. te Heesen, S., Janetzky, B., Lehle, L., and Aebi, M. (1992) The yeast WBP1 is essential for oligosaccharyl transferase activity *in vivo* and *in vitro*. *EMBO J.* **11,** 2071–2075
33. Helenius, J., Ng, D. T., Marolda, C. L., Walter, P., Valvano, M. A., and Aebi, M. (2002) Translocation of lipid-linked oligosaccharides across the ER membrane requires Rft1 protein. *Nature* **415,** 447–450
34. de Hoon, M. J., Imoto, S., Nolan, J., and Miyano, S. (2004) Open source clustering software. *Bioinformatics* **20,** 1453–1454
35. Jinnelov, A., Ali, L., Tinti, M., Güther, M. L. S., and Ferguson, M. A. J. (2017) Single-subunit oligosaccharyltransferases of *Trypanosoma brucei* display different and predictable peptide acceptor specificities. *J. Biol. Chem.* **292,** 20328–20341
36. Manthri, S., Güther, M. L., Izquierdo, L., Acosta-Serrano, A., Ferguson, M. A. (2008) Deletion of the TbALG3 gene demonstrates site-specific *N*-glycosylation and *N*-glycan processing in *Trypanosoma brucei*. *Glycobiology* **18,** 367–383
37. Igura, M., Maita, N., Kamishikiryo, J., Yamada, M., Obita, T., Maenaka, K., and Kohda, D. (2008) Structure-guided identification of a new catalytic motif of oligosaccharyltransferase. *EMBO J.* **27,** 234–243
38. Lizak, C., Gerber, S., Numao, S., Aebi, M., and Locher, K. P. (2011) X-ray structure of a bacterial oligosaccharyltransferase. *Nature* **474,** 350–355
39. Schulz, B. L., and Aebi, M. (2009) Analysis of glycosylation site occupancy reveals a role for Ost3p and Ost6p in site-specific *N*-glycosylation efficiency. *Mol. Cell. Proteomics* **8,** 357–364
40. Xu, Y., Bailey, U. M., and Schulz, B. L. (2015) Automated measurement of site-specific *N*-glycosylation occupancy with SWATH-MS. *Proteomics* **15,** 2177–2186
41. Zacchi, L. F., and Schulz, B. (2016) SWATH-MS glycoproteomics reveals consequences of defects in the glycosylation machinery. *Mol. Cell. Proteomics* **15,** 2435–2447
42. Guthrie, C., and Fink, G. (1991) Guide to yeast genetics and molecular biology. *Methods Enzymol.* **194,** 1–863
43. Zufferey, R., Knauer, R., Burda, P., Stagljar, I., te Heesen, S., Lehle, L., and Aebi, M. (1995) STT3, a highly conserved protein required for yeast oligosaccharyl transferase activity *in vivo*. *EMBO J.* **14,** 4949–4960
44. Boeke, J. D., LaCroute, F., and Fink, G. R. (1984) A positive selection for mutants lacking orotidine-5′-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. *Mol. Gen. Genet.* **197,** 345–346
45. Mueller, S., Wahlander, A., Selevsek, N., Otto, C., Ngwa, E. M., Poljak, K., Frey, A. D., Aebi, M., and Gauss, R. (2015) Protein degradation corrects for imbalanced subunit stoichiometry in OST complex assembly. *Mol. Biol. Cell* **26,** 2596–2608
46. Wiśniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009) Universal sample preparation method for proteome analysis. *Nat. Methods* **6,** 359–362
47. MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., Kern, R., Tabb, D. L., Liebler, D. C., and MacCoss, M. J. (2010) Skyline: An open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26,** 966–968

Analysis of substrate specificity of *Trypanosoma brucei* oligosaccharyltransferases (OSTs) by functional expression of domain-swapped chimeras in yeast

Kristina Poljak, Jörg Breitling, Robert Gauss, George Rugarabamu, Mauro Pellanda and Markus Aebi

J. Biol. Chem. 2017, 292:20342-20352.
doi: 10.1074/jbc.M117.811133 originally published online October 17, 2017